

JFM



Patent
Attorney's Docket No. 0026-0072

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of)
Shioupyn Shen) Group Art Unit: 2171
Application No.: 10/808,326) Examiner: Unassigned
Filed: March 25, 2004)
For: DOCUMENT NEAR-DUPLICATE)
DETECTION)

INFORMATION DISCLOSURE STATEMENT
TRANSMITTAL LETTER

U.S. Patent and Trademark Office
Customer Service Window, Mail Stop Amendment
Randolph Building
401 Dulany Street
Alexandria, VA 22314

Sir:

Enclosed is an Information Disclosure Statement and accompanying form PTO-1449 for the above-identified patent application.

- No additional fee for submission of the IDS is required.
- The fee of \$180.00 as set forth in 37 C.F.R. § 1.17(p) is also enclosed.
- A certification under 37 C.F.R. § 1.97(e) is also enclosed.
- Charge \$ _ to Deposit Account No. 50-1070 for the fee due.
- A check in the amount of \$ _ is enclosed for the fee due.

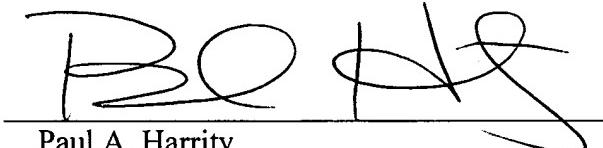
Information Disclosure Statement Transmittal Letter
Application Serial No. 10/808,326
Attorney's Docket No. 0026-0072
Page 2

The Commissioner is hereby authorized to charge any other appropriate fees that may be required by this paper that are not accounted for above, and to credit any overpayment, to Deposit Account No. 50-1070.

Respectfully submitted,

HARRITY & SNYDER, L.L.P.

By:


Paul A. Harrity
Reg. No. 39,574

11240 Waples Mill Road
Suite 300
Fairfax, Virginia 22030
(571) 432-0800

CUSTOMER NUMBER: 44989

Date: June 24, 2005



Patent
Attorney's Docket No. 0026-0072

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of)
Shioubyn Shen) Group Art Unit: 2171
Application No.: 10/808,326) Examiner: Unassigned
Filed: March 25, 2004)
For: DOCUMENT NEAR-DUPLICATE)
DETECTION)

INFORMATION DISCLOSURE STATEMENT UNDER 37 C.F.R. § 1.97(b)

U.S. Patent and Trademark Office
Customer Service Window, Mail Stop Amendment
Randolph Building
401 Dulany Street
Alexandria, VA 22314

Sir:

Pursuant to 37 C.F.R. §§ 1.56 and 1.97(b), Applicant(s) bring to the attention of the Examiner the documents listed on the attached PTO 1449. This Information Disclosure Statement is being filed before the mailing date of a first Office Action in the above-referenced application. As such, no certification or fee is required. Copies of the non-U.S. patent documents are attached.

Applicant(s) respectfully request(s) that the Examiner consider the listed documents and indicate that they were considered by making appropriate notations on the attached form.

Information Disclosure Statement Under 37 C.F.R. § 1.97(b)
Application Serial No. 10/808,326
Attorney's Docket No. 0026-0072
Page 2

If any copending application(s) is/are cited on the attached PTO 1449, the Examiner's attention is directed to the foregoing application(s) in compliance with § 2001.06(b) of the Manual of Patent Examining Procedure. By identifying the copending application(s), the assignee and/or applicant of the application(s) do not waive confidentiality of the application(s). Accordingly, the U.S. Patent and Trademark Office is requested to maintain the confidentiality of the copending application(s) under 35 U.S.C. § 122.

This submission does not represent that a search has been made and does not constitute an admission that each or all of the listed documents are material or constitute "prior art." If the Examiner applies any of the documents as prior art against any claim in the application and Applicant(s) determine(s) that the cited document(s) do not constitute "prior art" under United States law, Applicant(s) reserve(s) the right to present to the Office the relevant facts and law regarding the appropriate status of such documents.

Applicant(s) further reserve(s) the right to take appropriate action to establish the patentability of the disclosed invention over the listed documents, should one or more of the documents be applied against the claims of the present application.

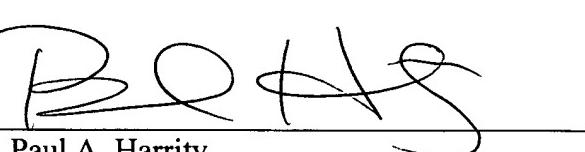
Information Disclosure Statement Under 37 C.F.R. § 1.97(b)
Application Serial No. 10/808,326
Attorney's Docket No. 0026-0072
Page 3

If there is any fee due in connection with the filing of this Statement, please charge the fee to our Deposit Account No. 50-1070.

Respectfully submitted,

HARRITY & SNYDER, L.L.P.

By:



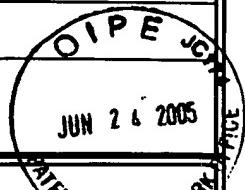
Paul A. Harrity
Reg. No. 39,574

11240 Waples Mill Road
Suite 300
Fairfax, Virginia 22030
(571) 432-0800

CUSTOMER NUMBER: 44989

Date: June 24, 2005

INFORMATION DISCLOSURE CITATION PTO-1449		CUSTOMER NUMBER 44989	ATTORNEY'S DKT NO. 0026-0072	APPLICATION NO. 10/808,326
			APPLICANT(S) Shioupyn Shen	
			FILING DATE March 25, 2004	GROUP 2171

**U.S. PATENT DOCUMENTS**

EXAMINER'S INITIALS	PATENT NO.	DATE	NAME	CLASS	SUBCLASS	FILING DATE
	6,658,423 B1	12/2/2003	Pugh et al.	707	102	

FOREIGN PATENT DOCUMENTS

EXAMINER'S INITIALS	PATENT NO.	DATE	COUNTRY	CLASS	SUBCLASS	Translation	
						Yes	No

OTHER DOCUMENTS (Including Author, Title, Date, Pertinent Pages, Etc.)

Arvind Arasu et al.; Extracting Structured Data from Web Pages; Proceedings of the ACM SIGMOD 2003, 2003; 30 pages.
Brenda S. Baker; A Theory of Parameterized Pattern Matching: Algorithms and Applications (Extended Abstract); 25 th ACM STOC 1993; 1993; pages 71-80
Brenda S. Baker; On Finding Duplication and Near-Duplication in Large Software Systems; Proceedings of the 2 nd Working Conference on Reverse Engineering; 1995; 10 pages.
Andrei Z. Broder; On the Resemblance and Containment of Documents; Proceedings of the Compression and Complexity of Sequences; 1997; pages 1-9.
Krishna Bharat et al.; Mirror, Mirror on the Web: A Study of Host Pairs of Replicated Content; Proceedings of the 8 th International Conference on World Wide Web (WWW 1999); 17 pages.
Krishna Bharat et al.; A Comparison of Techniques to Find Mirrored Hosts on the WWW; Journal of the American Society for Information Science; 2000; 11 pages.
Krishna Bharat; The Connectivity Server: Fast Access to Linkage Information on the Web; Proceedings of the 7 th International Conference on the World Wide Web; 1998; 13 pages.
Andrei Z. Broder et al.; Min-Wise Independent Permutations; Proceedings of STOC; 1998; pages 630-659.
Sergey Brin et al.; Copy Detection Mechanisms for Digital Documents; Proceedings of the ACM SIGMOD Annual Conference, 1995; 12 pages.
Andrei Z. Broder et al.; Syntactic Clustering of the Web; Proceedings of WWW6, 1997; 13 pages.
James W. Cooper et al.; Detecting Similar Documents Using Salient Terms; Proceedings of the CIKM 2002; November, 2002; pages 245-251.
Edith Cohen et al.; Finding Interesting Associations without Support Pruning; Proceedings of the 16 th ICDE; 2000; 12 pages.
Abdur Chowdhury et al.; Collection Statistics for Fast Duplicate Document Detection; ACM Transactions on Information Systems; Vol. 20, No. 2; April 2002; pages 171-191.
Zhiyuan Chen et al.; Selectively Estimation for Boolean Queries; Proceedings of PODS 2000; 2000; 10 pages.
Jack G. Conrad et al.; Constructing a Text Corpus for Inexact Duplicate Detection; SIGIR 2004; July, 2004; pages 582-583.
Scott Deerwester et al.; Indexing by Latent Semantic Analysis; Journal of the American Society for Information Science; 1990; 34 pages.

EXAMINER	DATE CONSIDERED

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant(s).

INFORMATION DISCLOSURE CITATION PTO-1449		CUSTOMER NUMBER 44989	ATTORNEY's DKT No. 0026-0072	APPLICATION No. 10/808,326
APPLICANT(S) Shioupyn Shen				
FILING DATE March 25, 2004	GROUP 2171			

U.S. PATENT DOCUMENTS

EXAMINER'S INITIALS	PATENT NO.	DATE	NAME	CLASS	SUBCLASS	FILING DATE

FOREIGN PATENT DOCUMENTS

EXAMINER'S INITIALS	PATENT NO.	DATE	COUNTRY	CLASS	SUBCLASS	Translation
						Yes No

OTHER DOCUMENTS (Including Author, Title, Date, Pertinent Pages, Etc.)

	Jeffrey Dean et al.; Finding Related Pages in the World Wide Web; Proceedings of the Eighth International World Wide Web Conference; 1999; pages 1-15.
	Aristides Gionis et al.; Efficient and Tunable Similar Set Retrieval; Proceedings of SIGMOD 2001; 2001; 12 pages.
	Taher H. Haveliwala et al.; Scalable Techniques for Clustering the Web; Proceedings of the 3 rd International Workshop on the Web and Databases (WebDB 2000); 2000; 6 pages.
	Taher H. Haveliwala et al.; Evaluating Strategies for Similarity Search on the Web; Proceedings of the 11 th International World Wide Web Conference; May, 2002; 11 pages.
	Khaled M. Hammouda et al.; Efficient Phrase-Based Document Indexing for Web Document Clustering; IEEE Transactions on Knowledge and Data Engineering; Vol. 16, No. 10, October 2004; pages 1279-1296.
	Timothy C. Hoad et al.; Methods for Identifying Versioned and Plagiarised Documents; Journal of the American Society for Information Science and Technology; 2003; pages 1-18.
	Sachindra Joshi et al.; A Bag of Paths Model for Measuring Structural Similarity in Web Documents; Proceedings of the 9 th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD 2003); August, 2003; pages 577-582.
	Jon M. Kleinberg; Authoritative Sources in a Hyperlinked Environment; Journal of the ACM; 1999; 34 pages.
	Aleksander Kotcz et al.; Improved Robustness of Signature-Based Near-Replica Detection via Lexicon Randomization; SIGKDD 2004; August, 2004; 6 pages.
	Ravi Kumar et al.; Trawling the Web for Emerging Cyber-Communities; Computer Networks: The International Journal of Computer and Telecommunications Networks; 1999; 21 pages.
	Udi Manber; Finding Similar Files in a Large File System; Proceedings of the 1994 USENIX Conference; 1994; 11 pages.
	Athicha Muthitacharoen et al.; A Low-Bandwidth Network File System; Proceedings of the 18 th ACM Symposium on Operating System Principles (SOSP 2001); 2001; 14 pages.
	Sean Quinlan et al.; Venti: A new Approach to Archival Storage; First USENIX Conference on File and Storage Technologies; 2002; 13 pages.

EXAMINER	DATE CONSIDERED
----------	-----------------

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant(s).

INFORMATION DISCLOSURE CITATION PTO-1449		CUSTOMER NUMBER 44989	ATTORNEY's DKT No. 0026-0072	APPLICATION NO. 10/808,326
			APPLICANT(S) Shioupyn Shen	
			FILING DATE March 25, 2004	GROUP 2171

U.S. PATENT DOCUMENTS

EXAMINER'S INITIALS	PATENT NO.	DATE	NAME	CLASS	SUBCLASS	FILING DATE

FOREIGN PATENT DOCUMENTS

EXAMINER'S INITIALS	PATENT NO.	DATE	COUNTRY	CLASS	SUBCLASS	Translation	
						Yes	No

OTHER DOCUMENTS (Including Author, Title, Date, Pertinent Pages, Etc.)

	Narayanan Shivakumar et al.; SCAM: A Copy Detection Mechanism for Digital Documents; Proceedings of the 2 nd International Conference on Theory and Practice of Digital Libraries; 1995; 9 pages.
	Saul Schleimer; Winnowing: Local Algorithms for Document Fingerprinting; Proceedings of SIGMOD 2003; June 9-12, 2003; 10 pages.
	Moses S. Charikar; Similarity Estimation Techniques from Rounding Algorithms; STOC 2002; May 19-21, 2002; 9 pages.
	Gautam Pant et al.; Crawling the Web; 2003; 25 pages.

EXAMINER	DATE CONSIDERED
----------	-----------------

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant(s).